

DIGITEK II

A Program For
Test Scoring and Analysis

User's Manual

IU BLOOMINGTON
EVALUATION SERVICES AND TESTING

INDIANA UNIVERSITY
BLOOMINGTON, INDIANA

Revised: June, 2007

TABLE OF CONTENTS

1	Introduction and logistics	2
1.1	Submitting sheets.....	2
1.2	Turn-around time.....	2
2	Student answer sheets	3
3	Filling out the key sheet(s)	3
3.1	The simplest possible key sheet.....	3
3.2	Allowing multiple correct answers	4
3.3	Counting all answers as correct	4
3.4	Skipping a question.....	5
4	Using subscales to group test questions	5
5	Using “alternate forms” to have multiple versions of a test.....	5
6	Interpreting the Digitek output.....	7
6.1	First page summary information section	8
6.1.1	The run overview.....	8
6.1.2	The scoring overview.....	9
6.1.3	Indices of reliability	9
6.2	List of keys and student responses	10
6.3	Distribution of total scores	11
6.4	Names, scores, ranks, and T-scores	11
6.5	Item statistics.....	12
7	Using Item Analysis results	15
8	Options for Digitek output.....	16
8.1	Default Digitek output options	16
8.2	Changing the sorting options	16
8.3	Changing which sections of the report are printed.....	17
8.4	Getting electronic versions of data.....	17
9	APPENDIX A: Scanned File Format.....	18

LIST OF FIGURES

Figure 1.	BEST scanning room location and hours	2
Figure 2.	Turn-around time for BEST <i>Digitek</i> submissions.....	2
Figure 3.	General Purpose answer sheet fields filled out in the simplest possible key sheet	4
Figure 4.	Fields in a secondary key sheet.	4
Figure 5.	Guidelines for using subscales	5
Figure 6.	Alternate Forms ordering. Examples can be seen in Figure 7 and Figure 8.	6
Figure 7.	Alternate Forms of a Ten-Item Test.....	7
Figure 8.	Alternate Forms of a Nine-Item Test	7
Figure 9.	Brief description of each Digitek data section.....	8
Figure 10.	The run overview section on the first page.....	8
Figure 11.	The scoring overview section on the first page	9
Figure 12.	Indices of reliability on the first page	9
Figure 13.	Keyed answers shown in the list of keys and student responses.....	10
Figure 14.	Student responses	11
Figure 15.	Frequency distribution of scores.....	11
Figure 16.	List of student names, scores, and ranks	12
Figure 17.	Part of a sample item analysis section	14
Figure 18.	Sorting options available.....	16

1 Introduction and logistics

Many instructors choose to use standard “bubble sheets” for quizzes and tests. *Digitek* is the test scoring service provided by The IU Bloomington Evaluation Services and Testing (BEST) office, and includes test scanning, scoring, and item analysis.

To use this process, instructors must use the General Purpose Answer Sheets (GPAS) to collect student data, fill out one or more GPAS sheets to act as correct answer keys, and submit the stack of key(s) and answer sheets to the BEST scanning room.

Location: Scanning Room: Franklin Hall M005 (855-3357)

BEST Main Office: Franklin Hall 014 (855-1595)

Hours: Monday – Friday, 8a.m. – 5p.m.

Exceptions posted on scanning room door and our web site:

<http://www.indiana.edu/~best>

Figure 1. BEST scanning room location and hours

1.1 Submitting sheets

After administering the test, the instructor should check over the answer sheets to be certain that all marks have been made properly: in pencil, dark enough and with circles filled in completely. Bring the answer sheets and the KEY to the BEST scoring room, Franklin Hall M005. Arrange all answer sheets in the same direction with side one face up. The key(s) should be on top of the sheets. The scanner will reject sheets that are backwards or upside down. Alphabetizing by name is **not** necessary.

1.2 Turn-around time

The scanning of answer sheets and subsequent test scoring and item analysis is performed on a first-come first-served basis. The time required depends on the week of the semester and time of day. The volume of scanning is heavy during mid- and end-of-semester. You should anticipate longer turn around time during these weeks. Also, the volume of tests is noticeably heavier in the afternoons.

To ensure that tests for all faculty are processed as expeditiously as possible, we do not “scan-while-you-wait.” Instead, we ask that you drop off your tests and return for pick-up at the time given by the scanning technician. The actual turn-around time depends on volume of work, but generally falls within the ranges listed below.

Regular semester	1–4 working hours
Finals week	2–6 working hours

Figure 2. Turn-around time for BEST *Digitek* submissions

2 Student answer sheets

The NCS General Purpose Answer Sheet (Forms #4521 or #214566) must be used by faculty for the Digitek scoring of multiple choice tests. This answer sheet may be requisitioned by departmental offices from Printing Services (call 5-6072) in packages of 500 sheets. This answer sheet is available in blue or red. Faculty who use alternate forms of their classroom tests often use both colors for ease in sorting answer sheets.

The answer sheet is designed for multiple choice tests with up to five options per item. These answer sheets can also be used for true-false tests by specifying that A=true and B=false. Photocopies of the answer sheet **cannot** be machine scored.

Students should indicate their name and/or student ID by filling in the bubbles in the appropriate columns of the answer sheet. Additionally, the section number can be put into the first five columns of the “Special Codes” section (columns K-O) if the instructor wishes to have this information. The “Sex,” “Birth Date,” “Educ,” and “Grade” columns are ignored by the scoring program.

Examinees should be instructed to use a #2 PENCIL for making all responses on the answer sheet. The scanner does not recognize ink of any kind. Students should also be instructed to blacken in only one answer per item. Items with multiple responses will be scored as incorrect.

The accuracy of machine scoring is highly dependent on the physical condition of the sheets. Sheets to be scanned should be kept clean, flat, and should not be stapled or have any holes punched in them. Sheets that have been wet may not scan properly even after they dry.

3 Filling out the key sheet(s)

An instructor must fill out one or more “key sheets” with the correct answers filled in. The key sheet is the same form as the student answer sheet (NCS General Purpose Answer Sheet, or GPAS). The name field of the key sheet must contain an identifying mark (‘XXXX’) followed by the instructor’s last name. The correct answer for each question must be indicated in the appropriate item field.

3.1 The simplest possible key sheet

The simplest possible key sheet contains the instructor’s last name preceded by four ‘X’ characters and exactly one correct answer for each question. If desired, the instructor may enter the test date in the Birth Date field and the course section number in the Special Codes field. If so, these will be displayed on the Digitek report as “Test Date” and “Codes,” respectively; however, they are not required.

Note: It is possible to re-use the same key sheet several times; however, the instructor should carefully check the sheet’s physical condition. Excessive dirt or small tears on the edges indicate that the key should be replaced.

Columns	Description	Example
The Name section	Four 'X' characters followed by the instructor's name	XXXXSmith
Items 1-200	For each item, fill in the circle corresponding to the correct answer with a #2 pencil	Item 1: D; Item 2: B; etc.

Figure 3. General Purpose answer sheet fields filled out in the simplest possible key sheet

3.2 Allowing multiple correct answers

While multiple choice items are ordinarily not intentionally constructed with more than one correct answer per item, this option is quite convenient for rescoreing tests which have been re-keyed following feedback from item statistics and from examinees.

You cannot key more than one correct answer for an item on a single key sheet. A multiply marked answer on a single sheet (e.g., filling in both the 'A' and 'C' circles for a single item) is unreadable by the scanner, and will result in *all* answers for that question being counted as correct.

Instead, you must include a second (or third or even fourth) key sheet, which includes only the 'XXXX' at the beginning of the name field and the additional correct answer(s). For this, you fill in circles for *only* those items that have multiple answers.

This means that your second key sheet may very well consist of only 'XXXX' in the name field and a single filled in circle for an item (e.g., 'D' for item 3), assuming you have only one question for which you want to allow multiple answers.

Note that duplicate answers on multiple key sheets are an error which may invalidate the scoring.

Columns	Description	Example
The Name section	Four 'X' characters, indicating a key	XXXX
Items 1-200	For each item with more than one correct response, fill in the circle corresponding to the correct secondary answer with a #2 pencil. Leave all other items blank.	Item 2: B Item 11: D

Figure 4. Fields in a secondary key sheet.

3.3 Counting all answers as correct

To count all answers to a question as correct, regardless of what answer the student may have entered, simply fill in all answers for that item on a key sheet. This is recognized by the scanner as an error and results in any student answer being counted correct.

3.4 Skipping a question

If you decide to “throw out” a question, simply don’t include any valid answers for that item on any of your key sheets. The question will be skipped, and thus eliminated from the total score and any statistics being generated. Student responses will not be considered either correct or incorrect.

4 Using subscales to group test questions

When an instructor gives a test that addresses two or more content areas, he or she may wish to note how students perform on the questions dedicated to each content area, or *subscale*.

To decompose your test into subscales, you use multiple key sheets, one for each subscale, that conform to the following guidelines.

1. Each key sheet must contain the leading ‘XXXX’ in the name field
2. The first key sheet should contain all the “normal” key sheet data (XXXXlastname) as well as answers for only the first subscale.
3. The second key sheet should contain answers for only the second subscale, and so on for all your subscales.
4. Additional key sheets (up to a maximum of four) can be used to allow multiple answers (see section 3.2, above).
5. No key sheet can contain both a new subscale item (items seen for the first time) and secondary answers (items seen for the second time).
6. It is an error to give an item the *same* correct answer on multiple key sheets; this may invalidate the scoring.

Figure 5. Guidelines for using subscales

In the most common case, the instructor will simply have one key sheet for each subscale.

The resulting analysis will show how each student did on each subscale.

5 Using “alternate forms” to have multiple versions of a test

Normally, if different versions of the same test are to be scored, each version must be scored separately with its own key. There is, however, a special option of the Digitek program that allows up to four versions of the same test to be scored together. This is called ALTERNATE FORMS.

Note that you must notify the scanning technician if you wish your tests to be scored using alternate forms!

There are four *forms*, or orderings of the items, available, as indicated in Figure 6. All forms are scored from the same key sheet. Digitek will rearrange student answers to

correspond to the “Form 1” order before scoring. Thus, the list of responses on the Digitek printout will be in the rearranged order.

To take advantage of this feature, the instructor must first prepare test booklets that arrange the questions into at least two of the forms listed in Figure 6. Second, *Alternate Forms* scoring requires that each student's answer sheet must indicate which form of the test the student took. The number of the form is marked in column P of the SPECIAL CODES box on the student's answer sheet. Thus, if there is a 1, 2, 3, or 4 in that position on a student sheet, it will be assumed that the order of questions on the test corresponds to the designated order for the desired alternate form. The students' answers are then rearranged in the order of Form 1 before the tests are scored.

Note: If the P code position on a student answer sheet is blank or contains anything other than a 1, 2, 3, or 4 it will be assumed that the test is Form 1.

Form 1	The original test.
Form 2	The items of the original test (Form 1) appear in reverse order.
Form 3, even # of items	The original test is divided in half and the last half is presented followed by the first half.
Form 4, even # of items	The original test is divided in half and the items of the first half of the original test appear in reverse order followed by the items of the second half of the original in reverse order.
Forms 3 and 4, odd # of items	The ordering is as described for Forms 3 and 4 above, except that the split is made just before the middle item, so that the middle item of the original test will always be grouped with the original second part. <i>Example:</i> In a 5-item quiz, group A consists of items 1 and 2; group B consists of items 3, 4, and 5.

Figure 6. Alternate Forms ordering. Examples can be seen in Figure 7 and Figure 8.

Instructors should use the examples below to insure that their arrangement of test forms is valid. Any error in the rearrangement of the test questions will invalidate the results. If this occurs, each test form will have to be run separately, with a separate key for each.

FORM 1 (Original Test)	FORM 2 Reversed	FORM 3 Split	FORM 4 Reverse/ Split
1. ITEM 1	1. ITEM 10	1. ITEM 6	1. ITEM 5
2. ITEM 2	2. ITEM 9	2. ITEM 7	2. ITEM 4
3. ITEM 3	3. ITEM 8	3. ITEM 8	3. ITEM 3
4. ITEM 4	4. ITEM 7	4. ITEM 9	4. ITEM 2
5. ITEM 5	5. ITEM 6	5. ITEM 10	5. ITEM 1
6. ITEM 6	6. ITEM 5	6. ITEM 1	6. ITEM 10
7. ITEM 7	7. ITEM 4	7. ITEM 2	7. ITEM 9
8. ITEM 8	8. ITEM 3	8. ITEM 3	8. ITEM 8
9. ITEM 9	9. ITEM 2	9. ITEM 4	9. ITEM 7
10. ITEM 10	10. ITEM 1	10. ITEM 5	10. ITEM 6

Figure 7. Alternate Forms of a Ten-Item Test

FORM 1 (Original Test)	FORM 2 Reverse	FORM 3 Split	FORM 4 Reverse/Split
1. ITEM 1	1. ITEM 9	1. ITEM 5	1. ITEM 4
2. ITEM 2	2. ITEM 8	2. ITEM 6	2. ITEM 3
3. ITEM 3	3. ITEM 7	3. ITEM 7	3. ITEM 2
4. ITEM 4	4. ITEM 6	4. ITEM 8	4. ITEM 1
5. ITEM 5	5. ITEM 5	5. ITEM 9	5. ITEM 9
6. ITEM 6	6. ITEM 4	6. ITEM 1	6. ITEM 8
7. ITEM 7	7. ITEM 3	7. ITEM 2	7. ITEM 7
8. ITEM 8	8. ITEM 2	8. ITEM 3	8. ITEM 6
9. ITEM 9	9. ITEM 1	9. ITEM 4	9. ITEM 5

Figure 8. Alternate Forms of a Nine-Item Test

6 Interpreting the Digitek output

The Digitek output has five main sections:

First page summary information , which includes the number of sheets read, mean and standard deviation for the total scores, and two indices of reliability.
List of keys and student responses , a complete list of the answers in each key and then the answers each student gave as a long list (e.g., AABACCAACADDCCDA). This information is, of course, available from the actual test sheet as well, which can be found by matching the sheet serial number in this section with the actual test sheet.
Distribution of total scores as a table, showing how many students received each total score.
Names, scores, and ranks gives the name, scores for each subscale, the total score, the <i>T-Score</i> ranking students relative to each other, and

<p>a count of how many questions they left blank and multiply marked (filling in more than one bubble, resulting in an error). This section also contains the sheet number, for easily pulling out a particular student's answer sheet.</p>
<p>Item statistics provides an analysis of each item, showing what percentage of students got it correct, what percentage gave each possible answer, how the question as a whole and each answer correlate with the total test score, how high-performing and low-performing students performed on the question, and how many students skipped or multiply marked this question.</p>

Figure 9. Brief description of each Digitek data section

With the exception of the summary information, which is always provided, any or all of these sections can be excluded by asking the operator to do so.

Each of these sections is explained below, along with a brief look at how to interpret the data they contain.

Note: For purposes of illustration, the following samples are drawn from more than one test; thus, the data may not be consistent from one section to the next.

6.1 First page summary information section

The summary information provided on the first page contains three small tables showing summary information and reliability data for the test run.

6.1.1 The run overview

RUN OVERVIEW		
Alternate forms not enabled		
229	Total Sheets Read	15 Test Items (1-15)
1	Key Sheet(s)	0 Items with all responses correct
4	Blank sheets detected	0 Items with more than one correct
4	Sheets rejected	
224	Examinees	

Figure 10. The run overview section on the first page

The **run overview** shows whether alternate forms were used, how many sheets were read, identified as keys, and possibly rejected, and which items are included on the test. You can also see how many items had more than one answer and how many items had all answers counted as correct (i.e., items with multiple marks, making them invalid).

6.1.2 The scoring overview

SCORING OVERVIEW	
Maximum Score	= 15
Minimum Score	= 5
Mean Score	= 13.38
Median Score	= 14
Standard Deviation	= 1.46

Figure 11. The scoring overview section on the first page

The **scoring overview** gives basic information about the total scores: minimum, maximum, mean, standard deviation, and median. This sample test didn't differentiate very well between students — nearly all did very well.

6.1.3 Indices of reliability

INDICES OF RELIABILITY	
Kuder-Richardson Reliability	= 0.48
Standard Error Of Measurement	= 1.05
Spearman-Brown Reliability	= 0.46
Standard Error Of Measurement	= 1.07

Figure 12. Indices of reliability on the first page

The **indices of reliability** show the value and standard error of measurement of two indices of test reliability: Kuder-Richardson and Spearman-Brown. Although calculated differently, both are indicators of the internal consistency of the test or the extent to which parts of the test are working together to measure the same underlying construct. Reliability indices range from 0 to 1.00. The higher the index, the more consistently and accurately the test is measuring.

Note that some tests, by design, will have low values on internal reliability. A test that covers two chapters of very divergent material, for example, will be measuring two very different sets of concepts and constructs and will therefore be expected to have low reliability. In these cases, consider using subscales (see section 4) to get a better idea of how students are doing on each content area.

Even when a single set of constructs is being tested, no test is perfectly reliable. The **standard error of measurement**, an index based on the reliability of the test, is used to predict the range of fluctuation likely to occur in a single individual's score as a result of chance errors. We can reasonably expect that about two-thirds of the time the student's obtained score on the test would be within \pm one standard error of measurement of his/her "true" score (the score the individual would get if there were no error in the measurement). For example, if a student has a test score of 20 and the standard error is 2, we can infer that the student's "true" score probably lies

somewhere between 18 and 22. There is one chance in six that it is lower than 18 and one chance in six that it is higher than 22.

On this sample data, we can see that out of 15 possible points, a student's "true score" is likely within ± 1 . Note the effect on standard grading: Suppose Mike got a $15 = 100\%$ (A+) and Bill got a $13 = 86\%$ (B). Based on this test only, there's a pretty good chance (1 in 3) that they actually know the material equally well and would, if they were given lots of chances to show it, end up with equivalent scores.

When test reliability is very high (near 1.00) the standard error of measurement is very small and we can have more confidence in the accuracy of the test scores.

Conversely, when reliability is low, the standard error of measurement will be large.

Because of test error, an instructor should be wary of attributing significance to minor differences in students' test scores. If test reliability is low, examinees must be many raw score points apart in order to legitimately conclude that they have significantly different amounts of knowledge. If test reliability is high, however, finer discrimination can be made among students.

6.2 List of keys and student responses

This section of the item analysis first lists the keyed answers to each of the test items. Next the printout shows a sorted list of the examinees along with their responses to each of the test items. If questions arise about the scoring of a particular item or items, the instructor has a record of the examinees' actual responses.

Figure 13 shows that two key sheets were used in scoring this test. An asterisk in place of the third keyed answer on Key Sheet 1 indicates that the instructor multiply marked this item. All valid responses to this item have been counted correct in scoring. Item 8 was omitted in scoring because it was not keyed on either key sheet. Items 9 and 30 have two correct answers, as indicated on Key Sheet 2.

	ITEM #s	KEYED ANSWERS
KEY 1	1- 30	D D * C C C A . D A A D B B D D A D B C A A A A B B B B A B
KEY 2	1- 30 C

Figure 13. Keyed answers shown in the list of keys and student responses

The sorting in this example is alphabetical by examinee's last name. One examinee neglected to fill in his/her name; thus his/her responses appear at the top of the list. Similarly, one student filled in two bubbles for the first letter, resulting in an asterisk that also sorts earlier than the letter 'A'. If you cannot find a student in the alphabetical listing, look at the top of the list for the record.

The *Sheet Number* is printed on the edge of each sheet as it goes through the scanner, in case you want to examine a student's actual answer sheet to find, e.g., off-by-one errors.

The responses are listed in groups of ten. Several periods and asterisks among the examinees' responses indicate items omitted and multiply marked, respectively.

Examinee's Names	ID No.	Sheet No.	Items	Examinees' Responses
	351520688	000104	1- 30	DDDCBCADDA CBCCDBCADC AAAABABBAB
*ILLENERG JIM	315541654	000105	1- 30	DDDCCCADDA BBCDABBD .AAAB.BBAB
ABRAM RUSSELL	314665960	000195	1- 30	CDACCCADDA CACBDDBADB AABABBBB
AL HASSAN AWAD	305485128	000094	1- 30	*DBCCCBDDA ADBBDDADBC AAAABBBB
DEBASSI SALMAN	312500740	000024	1- 30	CDDCBCADCD CDCAABABAD ABAAAABBB
DRAKE EDNA	308566922	000228	1- 30	CDDCBCADCD AAACDAABDC A*ABABBA.
MILLER LOIS	313707879	000186	1- 30	B*DCCCADDA AAACDBCADC AAAABBBB
NAJEM MARY BETH	498646372	000114	1- 30	ADDCCCBDD CDBCDCBCC ABAAABBB
POLLACK BARRY	311747699	000227	1- 30	ADDCACADCD CDDAABCDB AAABBASAAB
THOMPSON CS	280460519	000090	1- 30	DDDCCBADDA ABCCDBABDB ABAABBBB
WALTON EVER	303725220	000027	1- 30	CD.CBDDDD CACBDDCDCC AABABBBAA.
	308641376	000032	1- 30	DDDCCCDAAD CBBCCDCBBC AAAABBBB

Figure 14. Student responses

6.3 Distribution of total scores

The frequency of total scores gives both frequency and cumulative frequency counts for each raw score encountered during the processing. In the example shown in Figure 15, you can see that the most common score was 14 and that only 20% of students scored below 13 out of 15.

Frequency Distribution			
Raw Score	Frequency	Cumulative Frequency	Cumulative Percentage
15	48	224	100%
14	77	176	78%
13	53	99	44%
12	22	46	20%
11	14	24	10%
10	7	10	4%
9	2	3	1%
5	1	1	0%

Maximum possible score = 15.00
 Mean score = 13.38
 Standard deviation = 1.46

Figure 15. Frequency distribution of scores

6.4 Names, scores, ranks, and T-scores

This section lists the total scores, subscores (if subscales were used), percentile rank, T-score, sheet number, and blank/multiple-mark counts for each student. Also listed are the ID number and codes (not shown here to conserve space).

Examinee's name	Total score	Raw subscores				Pct. Rank	T-score	Items Blank	Multi Marked	Sheet No
		1	2	3	4					
*BER BILL	49	16	33			8	35			000024
ANDERSON ED	61	21	41			82	58		1	000018
CHELS EMILY	53	16	37			22	43			000115
DECKARD JOHN	65	21	44			98	66			001094

Figure 16. List of student names, scores, and ranks

In addition to raw scores, each total raw score has been converted to a **percentile rank** and to a **T-score**. The percentile rank shows the percentage of the class scoring lower than a particular examinee. For example, Ed Anderson has a percentile rank of 82. This means that 82% of the scores in the distribution are below his score. Students with the same raw score receive the same percentile rank. [Note: the actual definition of percentile rank is the percentage of students strictly below that score plus $\frac{1}{2}$ of the students with the same score.]

T-scores indicate how many standard deviation units an examinee's score is above or below the mean. T-scores always have a mean of 50 and a standard deviation of 10, so any T-score is directly interpretable. A T-score of 50 indicates a raw score equal to the mean. A T-score of 40 indicates a raw score one standard deviation below the mean, while a T-score of 65 indicates a raw score 1.5 standard deviations above the mean.

Both percentile rank and T-scores describe test performance in terms of the examinee's relative position in the distribution of test scores. While percentile rank has the advantage of being easier to understand, it has the serious disadvantage of representing a scale where the percentile units are not equal on all parts of the scale. A percentile rank difference near the middle of the scale (e.g., 45% to 55%) represents a much smaller difference in test performance than the same percentile difference at the ends (e.g., 85% to 95%). T-scores, on the other hand, provide equal units that can be treated arithmetically. T-scores from several tests taken during a semester can thus be summed and averaged. Percentile ranks should not be summed.

6.5 Item statistics

The *item statistics* section gives basic difficulty and correlation data for each question, and each answer for each question, plus counts of the missing/multiple-marked answers (if any) and how the best and worst scorers did on each question, as explained below.

- **Item No.** simply gives the item number. Items that had no answer given on any key are skipped. If this is preceded by a double-plus (++) all answers were considered correct for this item because a key had multiple marks in this space.
- **Item R.** The *R* in this case is the point-biserial correlation between a correct answer on the item and the total score. *R* ranges from -1 to +1. A positive *R* indicates that students who answered this question correctly were likely to do well on the test, and students who answered this question incorrectly were likely to do poorly on the test as a whole. An *R* near zero indicates that this question didn't

discriminate between those who did well overall and those who did poorly overall.

Roughly speaking, then, the R value is a measure of how representative this question is of the rest of the test. An item with a high R value is a good predictor of final total score; a negative R value would lead one to incorrectly predict final total score.

A question with a very low difficulty (i.e., almost all students answered it correctly) will have an R near zero (as will a question almost no one gets correct) since it is unable to discriminate between high- and low-scoring students. Item 5 in Figure 17 is an example; all students got the item correct, so there is obviously no difference on this item between high- and low-scorers and the R-value is zero.

A negative R value is a warning sign: this indicates that students who performed poorly overall were *more* likely to answer this item correctly than the high-performing students. This item should be looked at carefully for revision, since either something about it is confusing the better-prepared students or (more likely) your key was filled out incorrectly. Item 21 in Figure 17 is such a question. The vast majority of students (96%) gave the incorrect answer “D.”

- **Percentage for response options A, B, C, D, E.** The “pct” values for each answer indicate what percentage of students who answered this question gave that particular answer.

Ideally, incorrect answers will be distributed evenly among the “distracters” (wrong answers), all of which will have a negative R value. A distracter that has a Pct near zero is not doing its job and might be a candidate for replacement.

- **R for response options A, B, C, D, E.** The R value of an individual answer indicates how good the choice of *that particular answer* is at predicting total score. The R value for the correct answer (indicated by an asterisk after the column) generally should be at least moderately high and positive, with R values for the incorrect answers being negative or near zero.

The Item R and the correct answer’s R will be identical, unless there is more than one correct answer for a question.

- **Top Third** is simply the percentage of students who scored in the top third of total scores on this test who got the item correct. **Bottom third** is the percentage of students who scored low overall who got this item correct. These columns act as a very rough (but still useful) alternative view of the correlation data (the R values). Obviously, Top third should always be higher than the Bottom third percentage, except for very easy items that nearly everyone got right.

See Section 7 for more information about interpreting the percentage and R values.

Response Categories												Percent correct	
Number													
ITEM NO.	Item	A	B	C	D	E	Top third	Bottom third					
		R Pct / R	R Pct / R	R Pct / R	R Pct / R	R Pct / R							
1	0.49	0 / 0.00	13 / -0.49	86 / 0.49 *	0 / 0.00	0 / 0.00	100	55					
2	0.42	6 / -0.14	82 / 0.42 *	0 / 0.00	10 / -0.41	0 / 0.00	100	66					
3	0.47	0 / 0.00	86 / 0.47 *	6 / -0.24	6 / -0.40	0 / 0.00	100	55					
4	-0.08	3 / 0.12	6 / -0.43	0 / 0.00	58 / -0.08 *	31 / 0.27	50	55					
5	0.00	0 / 0.00	100 / 0.00 *	0 / 0.00	0 / 0.00	0 / 0.00	100	88					
6	0.26	86 / 0.26 *	3 / -0.42	3 / -0.24	6 / 0.13	0 / 0.00	83	77					
7	0.51	0 / 0.00	93 / 0.51 *	6 / -0.51	0 / 0.00	0 / 0.00	100	77					
8	0.43	3 / -0.46	0 / 0.00	3 / -0.13	93 / 0.43 *	0 / 0.00	100	77					
9	0.35	0 / 0.00	0 / 0.00	6 / -0.35	93 / 0.35 *	0 / 0.00	100	77					
10	0.21	6 / -0.01	72 / 0.21 *	6 / 0.05	13 / -0.30	0 / 0.00	83	55					
11	0.48	93 / 0.48 *	6 / -0.48	0 / 0.00	0 / 0.00	0 / 0.00	100	77					
12	0.06	24 / 0.15	3 / -0.46	0 / 0.00	41 / -0.02	31 / 0.06 *	41	22					
13	0.46	0 / 0.00	96 / 0.46 *	3 / -0.46	0 / 0.00	0 / 0.00	100	88					
14	-0.05	0 / 0.00	86 / -0.05 *	13 / 0.05	0 / 0.00	0 / 0.00	83	77					
15	0.45	93 / 0.45 *	6 / -0.45	0 / 0.00	0 / 0.00	0 / 0.00	100	77					
16	0.39	3 / 0.01	82 / 0.39 *	0 / 0.00	13 / -0.43	0 / 0.00	100	66					
17	0.03	86 / 0.03 *	3 / 0.01	10 / -0.04	0 / 0.00	0 / 0.00	91	88					
18	0.46	0 / 0.00	96 / 0.46 *	3 / -0.46	0 / 0.00	0 / 0.00	100	88					
19	0.28	0 / 0.00	3 / -0.02	0 / 0.00	10 / -0.30	86 / 0.28 *	100	77					
20	0.50	0 / 0.00	27 / -0.33	3 / -0.46	68 / 0.50 *	0 / 0.00	83	44					
21	-0.20	0 / 0.00	3 / 0.20	0 / 0.00	96 / -0.20 *	0 / 0.00	91	100					
22	0.24	0 / 0.00	0 / 0.00	0 / 0.00	13 / -0.24	86 / 0.24 *	91	66					
23	0.35	0 / 0.00	0 / 0.00	93 / 0.35 *	6 / -0.35	0 / 0.00	100	88					
24	0.55	20 / -0.21	72 / 0.55 *	3 / -0.46	3 / -0.42	0 / 0.00	83	33					
25	0.32	3 / 0.01	6 / -0.11	0 / 0.00	86 / 0.32 *	3 / -0.46	91	77					
26	0.08	0 / 0.00	3 / 0.01	3 / -0.46	48 / 0.08 *	44 / 0.08	50	33					

Figure 17. Part of a sample item analysis section.

7 Using Item Analysis results

The Item Statistics and the Reliability information are useful in assessing both the adequacy of individual test items and the overall quality of the test. We should strive to construct tests that have a reliability index at least in the 0.70's, preferably higher. One way to increase the reliability of a test is to increase the length. *In general, the longer the test, the higher the reliability.* Having high quality items and careful administration of the test are other ways to increase the reliability.

The difficulty of a test item is indicated by the percentage of students who got the item right (i.e., the Pct value for the correct answer). Items that most students got wrong may mean the concept tested in the item was not covered well in class or that the item was not clearly written. Another cause of a very low correct percentage is the mis-keying of an item.

Items with very high percentages for the correct answer(s) are easy items. Having an easy item or two at the beginning of a test is recommended. But a test with many very easy items (or many very difficult items) will not discriminate well among examinees and consequently will have low reliability. Assuming that the goal is to distinguish between students who have mastered the course content and those who have not, items of moderate difficulty will spread the distribution of test scores much better. A rule of thumb suggests that multiple choice items have a difficulty level between 35% and 85%. Items falling outside this range are too difficult or too easy to discriminate well.

The percentages for the other answer options are also useful. Good distracters (incorrect answer options) should be selected by some students. Distracters with very few or no “takers” might be considered for replacement since they are not helping the test discriminate between students who have and have not mastered the information.

The item discrimination (R) is the index most useful in evaluating test items. A moderate (at least 0.30) or high positive R indicates that the item is discriminating among examinees in the same way that the total score is discriminating among them. A larger proportion of the more knowledgeable students than poor students (as indicated by total test score) got the item right. A discrimination index near zero indicates that the item is not discriminating in any systematic way relative to total score. The same number of low scorers as high scorers got the item right. An item with a negative discrimination index is working in a direction opposite to the total test scores: low total scorers tend to get the item correct; high total scorers tend to get the item wrong. This often indicates an item that has been mis-keyed.

Item statistics may be used to identify **potentially** poor items. A negative or zero discrimination index may indicate ambiguity, cues, mis-keying, or other technical defects to the item. Some items may need to be eliminated. Others with some revision and rewriting of the stem and/or options may become worthwhile items. When eliminating problem items from a test, one must be careful not to alter the representativeness of the sampling of learning outcomes and course content. Some

items with a low discrimination index or high difficulty level may need to be retained simply because they measure an important learning outcome.

8 Options for Digitek output

The default test scoring and item analysis setting is to provide all sections of the output; alternative options are described below.

8.1 Default Digitek output options

- Scores printed on answer sheets.
- Item analysis (printout containing examinee responses, frequency distribution, item statistics, and reliability coefficient).
- Examinee responses and scores printed in order as **sorted by name**.
- Items may have more than one right answer, if multiple key sheets indicate they should.
- Subscale scoring, if the key sheets are formatted as necessary
- Alternate test forms *disabled*; this option must be turned on by the scanning room operator.

8.2 Changing the sorting options

There are a number of sorting options which the operator can set for you on request.

Sorting options available
Do not sort; use the order in which the sheets were stacked
Sort alphabetically by name (default)
Sort by Identification Number
Sort by total score (descending), then by name
Sort by code, then by Identification Number
Sort by code, then by name

Figure 18. Sorting options available

Only one choice may be executed. Sorting is on the entire field. If the first column of a field is blank, the file will contain a space which will be included in the sort. The space character sorts **before** any letter or number.

Sorting on **codes** can be useful to those instructors who have their students enter their laboratory or discussion section numbers in the columns K-N of the “SPECIAL CODES” section of the answer sheet. This allows the instructor to group students by code, and then view the report sorted first by code, then by name or ID.

Note that the special sorting will be used only for the student scores section. The list of student responses is always sorted alphabetically.

8.3 Changing which sections of the report are printed

The Digitek printout provides five sections, described in section 6. The default is to provide each of these; however, you may ask the scanning technician to suppress sections you do not want or need.

8.4 Getting electronic versions of data

Please call and consult with the scanning room operator in order to find out which media are currently supported. You can get one or all of the following files returned to you along with your test sheets:

- Scanned data file: a fixed-field length ASCII file; the record format is given in Appendix A.
- Comma-delimited format: this file contains only the ID number, name, special codes fields, and the raw score(s).
- Excel file: A Microsoft Excel (.xls) workbook containing worksheets for both the student data (name, ID, code, total score, subscale-score, and answers) as well as the data provided by the item analysis section (see section 6.5 for details).

The ASCII files may be obtained in addition to the standard printout. The Excel file is provided in lieu of the printout.

9 APPENDIX A: Scanned File Format

RECORD FORMAT			
General Purpose Answer Sheet: 5-response			
NAME	LENGTH	START	END
Machine Code	40	1	40
Month	2	41	42
Date	2	43	44
Year	2	45	46
Sex	1	47	47
Grade	2	48	49
ID	10	50	59
Name	20	60	79
Special Code	6	80	85
Items 1-200	200	90	289
Total score	5	290	294
Sub-score 1	4	295	298
Sub-score 2	4	299	302
Sub-score 3	4	303	306
Sub-score 4	4	307	310